

## Artikkeli



# Tunteiden, teknologian ja journalismin ristipaineessa

## Verkkokomentoinnin moderointi suomalaisissa uutistoimituksissa

Uutisten verkkokomentoinnin kielteistä ilmapiiiriä ja asiattomia kommentteja pidetään vakavana, julkisen keskustelun laatua heikentävänä ongelmana. Keskustelujen moderointia pidetään yleisesti hyvänä tapana puuttua asiattomuuksiin, mutta sen nykyiset keinot koetaan riittämättömiksi, olipa kyse tietokoneohjelmista tai ihmisten tekemästä moderointityöstä. Artikkelissa esittelemme uutiskomenttien moderointia käsittelevää aikaisempaa tutkimusta sekä tarkastelemme aihetta empiirisesti analysoimalla viittä eri mediataloa edustavan suomalaisen media-alan ammattilaisen näkemyksiä ja kokemuksia verkkokomenttien moderoinnista, keskittyen erityisesti kommentteissa esiintyvien tunteiden rooliin. Analyysin pohjalta kommentteista ja niiden moderoinnista piiryy ristiriitainen kuva. Yhtäältä kommentteja arvostetaan yleisön osallistumisen muotona ja niiden tunnesisällön koetaan lisäävän keskustelun vetovoimaa. Toisaalta moderointi koetaan vaivalloisena ja journalistisesta perustyöstä erillisenä tehtävänä ja kommentit liian negatiivisina. Myös moderoinnin ja journalismin vaaliman ilmaisunvapauden suhde näyttäytyy jännitteisenä. Tietokoneavusteinen moderointi puolestaan koetaan hyödylliseksi, mutta sitä pidetään myös epätarkkana. Näitä havaintoja vasten pohdimme, millaista uutiskomenttien moderointi voisi tulevaisuudessa olla suomalaisella mediakentällä. Esitämme, että tehokkuutta tavoittelevan moderoinnin lisäksi käytössä olisi hyvä olla keinoja, jotka tarttuvat asiattomaan kommentointiin pehmeämmin kuin sisältöä poistamalla.

**AVAINSANAT:** uutiskomentointi, moderointi, verkkojournalismi, yleisöt, tunteet

Lukijakomentointi on vakiintunut osaksi sanomalehtien verkkosivustoja noin 2000-luvun ensimmäisen vuosikymmenen puolivälistä lähtien (Masullo Chen ja Pain 2017, 877). 1990-luvulla ja 2000-luvun alkuvuosina internetin hyödyntämiseen journalismissa kohdistui suuria odotuksia. Sen toivottiin demokratisoivan tiedonvälitystä ja julkisuuteen pääsyä sekä tarjoavan kansalaisille mahdollisuuksia osallistua niin poliittiseen päätöksentekoon kuin sanomalehtien jutuntekoon (Heinonen ym. 2000, 18).

Uutisten verkkokommentoinnin ajateltiin olevan eräs osallistuminen muoto, mutta arki on kuitenkin osoittautunut hyvin toisenlaiseksi. 2020-luvun alkuvuosiin tultaessa sekä populaarissa että tieteellisessä keskustelussa on esitetty enenevässä määrin huolestuneita puheenvuoroja verkon keskusteluilmapiiriin tulehtumisesta ja sitä seuraavasta jyrkästä poliittisesta polarisaatiosta (mm. Haapoja, Laaksonen ja Lampinen 2020, 2; Laaksonen ym. 2020, 2–3; Meltzer 2015, 86; Nikunen ja Pantti 2018, 72–73; Saresma ym. 2022, 12). Lisäksi keskustelun heikon laadun koetaan olevan haitallista mediatalon brändille ja uskottavuudelle, ja pahimmassa tapauksessa se aiheuttaa vahinkoa myös toimittajille ja lukijoille (Larsson 2018, 607–608; Masullo Chen ja Pain 2017, 879; Wolfgang, McConell ja Blackburn 2020, 925–926).

Verkkokommentoinnin varjopuolista on keskusteltu sekä tutkimuksen piirissä että laajemmassa julkisuudessa ainakin 1990-luvulta saakka (Gillespie 2018, 25–29). Verkkokeskustelun valtavirtaistumisen ja kaupallistumisen myötä nämä piirteet ovat tulleet osaksi 2000-luvun alkuvuosikymmenien huolipuhetta siitä, miten verkko vaikuttaa sosiaaliseen ja poliittiseen kanssakäymiseen (emt.). Esimerkiksi Suomessa on jo vuonna 2008 keskusteltu kiivaasti siitä, kuinka epätoivottua verkkokäyttäytymistä tulisi valvoa ja kitkeä (Haara 2012; Juntunen 2008; Pöyhtäri, Haara ja Raittila 2013).

Yhtenä ratkaisuna epäasiallisen kommentoinnin ongelmaan on sekä Suomessa että kansainvälisesti tarjottu tiukempaa moderointia niin journalismin kontekstissa kuin esimerkiksi sosiaalisessa mediassa. Yksi ratkaisuehdotuksista on ollut tietokoneavusteinen uutiskommenttien moderointi. Tietokoneavusteisella moderoinnilla tarkoitamme moderointikäytäntöjä, joissa tietokoneohjelma tekee joko ehdotuksia tai itsenäisiä päätöksiä moderoitavan sisällön suhteen. Tekoälyteknologian nopean kehityksen ansiosta tietokoneavusteiseen moderointiin on liittynyt niin Suomessa kuin kansainvälisesti paljon odotuksia (esim. Laaksonen ym. 2020, 4), jotka eivät kuitenkaan ole kaikin osin täyttyneet (Gillespie 2018, 181, 194–196; Gorwa, Binns ja Katzenbach 2020, 10–12; Laaksonen ym. 2020, 10). Suurimman osan moderointityöstä niin Suomessa kuin kansainvälisesti tekevätkin yhä ihmiset, joko palkattuina työntekijöinä tai vapaaehtoisina (Gillespie 2018; Roberts 2019).

Uutiskommentit ovat kulttuurisena ja sosiaalisena ilmiönä monitasoisia, ja niihin kietoutuu julkisen keskustelun problematiikan useita eri puolia. Kommentointi nojaa julkisen keskustelun ja demokratian ihanteisiin, kun lukijoilla on mahdollisuus kommentoida uutisia suoraan tiedotusvälineen omilla verkkosivuilla (Harju 2018, 53–54; Kangaspunta 2021, 74; Larsson 2018, 606). Samalla uutiskomentointi ja muu mediataloille tarjottu käyttäjien tuottama sisältö on usein tunteikasta tai jäsentymätöntä (Papacharissi 2015; Wahl-Jorgensen 2019), ja siten ne asettuvat usein agonistiseen suhteeseen (Mouffe 1999) järkeilevän julkisen keskustelun ihanteiden kanssa. Jäsentymätön keskustelu myös vaatii sivustojen ylläpitäjiltä moderointitoimia (Gillespie 2018). Heikot kokemukset keskustelujen moderoinnista ihmisten tai koneiden tekemänä kutsuvat kysymään, miten tietokoneohjelmistojen toiminta ja moderoinnin varsin binääriset periaatteet suhteutuvat uutiskommentoinnin sosiokulttuurisesti monikerroksiseen kontekstiin.

Komentoinnin tunteisuus vaikuttaa tuottavan jännitteitä journalistisessa kontekstissa käytävän julkisen keskustelun ihanteisiin. Sen tähden lähestymme artikkelissa uutisten verkkokommenttien moderointia tunteiden sekä edellä mainitun jännitteisen ja ristiriitaisen suhtautumisen näkökulmasta. Pyrimme hahmottamaan, miten verkkokommen-

toinnissa ilmentyvät tunteet liittyvät moderointiin. On tärkeää ymmärtää, miten tunteita ilmaistaan journalististen instituutioiden tuottamissa julkisen keskustelun tiloissa, ja miten ne ymmärretään yhtä aikaa ongelmallisiksi ja keskustelun autenttisuutta merkitseviksi sekä journalistissa teksteissä että kommenteissa. Lisäksi tarkastelemme artikkelissa, miten moderointityö suhteutuu muuhun journalistiseen työhön ja miten tunnenäkökulma tulee mahdollisesti näkyviin moderoinnin yhteydessä.

Artikkelissa nostamme esiin aikaisempaa tutkimuskirjallisuutta uutisten verkkokommentoinnista ja sen moderoinnista. Lisäksi tarkastelemme 11 suomalaisen media-alan ammattilaisen näkemyksiä ja kokemuksia uutiskomenttien moderoinnista ja sen tulevaisuuden näkymistä puolistrukturoitujen haastattelujen avulla. Analysoimme haastatteluaineistoa aineistolähtöisen laadullisen teema-analyysin keinoin ja reflektoimme tuloksiamme suhteessa aikaisempaan tutkimukseen.

Artikkeli vastaa seuraaviin tutkimuskysymyksiin:

1. Millaisena moderoinnin tutkimus näyttäytyy aikaisemman kirjallisuuden valossa?
2. Miten suomalaiset journalismin ammattilaiset hahmottavat moderoinnin suhteessa muuhun journalistiseen työhön?
3. Millaisia mahdollisia jännitteitä haastateltavien puheissa ilmenee koskien uutisten verkkokommentoinnin moderointia joko tietokoneiden avustamana tai ihmisten tekemänä?
4. Millaiseen rooliin verkkokomenttien tunneilmaisut asettuvat journalismin ammattilaisten puhuessa kommenttien moderoinnista?
5. Miten uutiskomenttien moderointia voisi kehittää tulevaisuudessa edellä esitettyjen kysymysten pohjalta?

Uutisten verkkokomentointia on Suomessa ja kansainvälisesti tutkittu enenevässä määrin. Suuri osa tutkimuksesta keskittyy kommentteihin osana journalistista julkisuutta (esim. Kangaspunta 2021, 86–95) ja journalistien näkemyksiin kommentteista (mm. Boberg ym. 2018; Frischlich ym. 2019; Heinonen 2011; Larsson 2018; Masullo Chen ja Pain 2017; Pöyhkäri, Haara ja Raittila 2013; Wolfgang 2018). Tässä artikkelissa esiteltävä tutkimus tuo keskusteluun uuden avauksen lähestymällä uutiskomentointia moderoinnin näkökulmasta.

Artikkelimme avaa näkymän kahteen perustavanlaatuisen ristiriitaan, jotka määrittävät uutiskomenttien moderointia tunteiden lisäksi journalistisen työn ja journalismi-instituution näkökulmasta. Ensimmäinen ristiriidoista liittyy julkisen keskustelun ihanteisiin ja ilmaisunvapauteen. Gillespietä (2018, 5) mukailen kaikkien alustojen on moderoitava sisältöään, jotta ne pysyvät käyttökelpoisina, mutta kuinka määritellä se, miten paljon tai vähän moderointia riittää? Toinen, konkreettisempi ristiriita tulee näkyviin, kun tarkastelemme journalistien kokemuksia moderoinnista. Moderointityö koetaan usein aikaa vieväksi ja siihen yhtenä ratkaisuna tarjottu tietokoneen apu toivotetaan tervetulleeksi. Tietokoneavusteiseen moderointiin liittyy kuitenkin suomalaisessa kontekstissa paljon toistaiseksi ratkaisemattomia teknologisia ja periaatteellisia ongelmia. Lisäksi esittelemme aikaisempaan tutkimukseemme nojaten ehdotuksia, joiden avulla näitä ristiriitoja olisi mahdollista liennyttää.

## Verkkokomentointi journalismin näkökulmasta

Aikaisemman tutkimuksen perusteella journalistien näkemyksen verkkokomentoinnista ovat kaksijakoisia. Yhtäältä kommentteja pidetään sekä yleisönosastokirjoittelun jatkeena, ja siten tärkeänä journalismin demokratiavaikutuksen kannalta, että keinona sitouttaa ja aktivoida journalismin yleisöjä (Graham ja Wright 2015; Heinonen 2011; Masullo Chen ja Pain 2017; Wolfgang, McConnell ja Blackburn 2020). Esimerkiksi Grahamin ja Wrightin (2015) tutkimuksessa brittilehti *The Guardianin* toimittajat kokivat, että lukijakommentit olivat niin tasokkaita, että niiden lukeminen ja keskusteluun osallistuminen tarjosi heille uusia näkökulmia omaan työhönsä (ks. myös Wright, Jackson ja Graham 2020). Masullo Chenin ja Painin (2017) yhdysvaltalaisen toimittajien suhdetta verkkokomentointiin selvittänyt tutkimus tuotti samanlaisia tuloksia: toimittajat kokivat kommentoijien kanssa keskustelun merkitykselliseksi sekä journalismi-instituution että oman työnsä kannalta.

Aiemman tutkimuksen mukaan toimittajat kokevat, että laadukkaan verkkokeskustelun ylläpito vaatii jatkuvaa työtä ja vie sen tuottamaan hyötyyn nähden paljon aikaa ja vaivaa. Esimerkiksi Heinosen (2011, 48) ja Reichin (2011, 102) laajaan kansainväliseen haastattelututkimukseen osallistuneet journalistit kokivat kommenttien moderointiin käytetyn ajan ja vaivannäön olevan pois journalismin ydintehtävistä. Samansuuntaisia tuloksia saatiin myös Wolfgangin, Blackburnin ja McConnellin (2020) sekä Winterlinin ja kumppaneiden (2020) tutkimuksissa, joista ensimmäisessä selvitettiin yhdysvaltalaisen ja toisessa saksalaisten toimittajien moderointikokemuksia. Toimittajien suhdetta uutisten verkkokomentointiin on tarkasteltu myös erilaisten vuorovaikutusta ja yhteisöllisyyttä edistävien journalismin muotojen kautta (esim. Ahva, Hautakangas ja Haara 2017; Heinonen 2011; Marchionni 2015; Masullo Chen ja Pain 2017; Reich 2011). Näissä tutkimuksissa toimittajien osallistumista kommentointiin ja/tai sen moderointiin pidetään osana toimittajan velvollisuutta yleisöään kohtaan. Osallistumista pidetään tärkeänä sekä journalismin yleisösuhteen että demokratiaa ja kansalaisaktiivisuutta edistävien vaikutusten tähden (Ahva, Hautakangas ja Haara 2017; Graham ja Wright 2015; Heinonen 2011; Masullo Chen ja Pain 2017; Reich 2011).

Yhteisöllisyyttä rakentavien journalismin muotojen lisäksi uutisten verkkokommentteja on tarkasteltu erilaisten julkisuusnäkökulmien kautta. Uutisten verkkokommentit mielletään usein journalististen tekstien kylkeen syntyväksi julkiseksi tilaksi, jossa kommentoijilla on mahdollisuus ajatusten vaihtoon ja yhteiseen järkeilyyn (Masullo Chen ja Pain 2017, 878; Ksaizek, Peer ja Lessard 2016, 504; Suh ym. 2018, 185). Esimerkiksi Kangaspunta (2021, 70), ja Wolfgang, McConell ja Blackburn (2020, 925) määrittävät tiedotusvälineiden verkkosivuilla uutisten ohessa julkaistavat kommentit osaksi tiedotusvälineiden tarjoamaa julkisen ajatuksenvaihdon tilaa. Suomalaisessa kontekstissa tätä sidosta vahvistaa Julkisen sanan neuvoston (JSN) vuonna 2011 julkaisema täydennys Journalistin ohjeisiin, jossa todetaan toimituksen olevan viime kädessä vastuussa esimerkiksi loukkaavien ja haitallisten kommenttien poistamisesta (Haara 2012, 54; Pöyhtäri, Haara ja Raittila 2013, 164, 177). Kuten Kangaspunta (2021, 48) huomauttaa, verkkokeskustelua lähestytään kirjallisuudessa yhtäaikaaisesti sekä *julkison* (engl. *public*, Ridell ja Pietilä 1998, 95) että *yleisön* (engl. *audience*) käsitteiden kautta riippuen siitä, tarkastellaanko verkkokeskusteluita yleisesti vai journalististen tekstien kommentointia. Molemmissa näkökulmissa korostuu vahva näkemys tai toive verkon julkisesta tilasta järkeilevän keskustelun paikkana habermasilaista julkisuuden

ihannetta mukaillen (Kangaspunta 2021, 63; Masullo Chen ja Pain 2017, 888; Papacharissi 2010, 119; Suh ym. 2018, 183).

Kun uutisten verkkokommentit mielletään osaksi journalismi-instituution tuottamaa julkisen keskustelun aluetta, kohdistuu kommentteissa käytävään keskusteluun myös erilaisia odotuksia ja toiveita (Boberg ym. 2018; Kangaspunta 2021; Larsson 2018; Winterlin ym. 2020). Näiden odotusten mukaan kommenttien toivotaan olevan ilmaisultaan asiallisia, hyvin muotoiltuja ja pitäytyvän keskustelun aiheessa sekä pidättäytyvän alatyylisistä ilmaisuista tai toisiin kommentoihin kohdistuvista loukkauksista (esim. Wolfgang, McConnell ja Blackburn 2020, 927). Suomalaisten mediatalojen lukijoille suunnatut kommentointiohjeet heijastelevat samanlaista ajattelutapaa.

Asiattoman ja vahvasti tunteikkaan kommentoinnin koetaan toimivan vastoin näitä odotuksia ja ihanteita, ja muun muassa siksi sitä pidetään ongelmallisena (Masullo Chen ja Pain 2017, 882; Larsson 2018, 607; Reich 2011; Wolfgang, Blackburn ja McConnell 2020, 435). Toisaalta Masullo Chen ja kumppanit (2019) huomauttavat, että kaikki vihainen puhe ei ole julkisen keskustelun kannalta huono asia, sillä ihmisillä tulee demokraattisissa yhteiskunnissa olla oikeus ilmaista negatiivisia tunteitaan julkisesti (ks. myös Mouffe 1999).

Julkisen keskustelun punnittua järkeilyä korostavien odotusten voi katsoa olevan ristiriidassa kommenttipalstojen arjen kanssa. Osittain tästä ristiriidasta syntyy tarve moderoida, toisin sanoen ohjata, muovata tai hallita kommentteissa käytävää keskustelua (Gillespie 2018). Ristiriita heijastuu myös siihen, miten toimittajat aikaisemman tutkimuksen perusteella suhtautuvat moderointiin. Heinosen (2011, 41, 51) sekä Pöyhtärin, Haaran ja Raittilan (2013, 169, 176, 179) mukaan suomalaiset toimittajat pitivät moderointia 2010-luvun alussa muulta työltä aikaa vievänä välttämättömänä pahana. Vaikka tutkimustulokset osoittavat, että toimittajien aktiivinen osallistuminen verkkokeskusteluihin tuo keskustelun lähemmäs sille osoitettuja ihanteita (esim. Graham ja Wright 2015; Masullo Chen ja Pain 2017), osa toimittajista ei pidä osallistumista aina vaivan arvoisena (Kangaspunta 2021; 88, Reich 2011, 102–106; Wolfgang 2018, 27).

## Uutisten verkkokomentointi ja tunteet

Uutisten verkkokomentoinnin asiattomuutta tai niiden heikkoa laatua perustellaan usein kommenttien tunteellisuudella (Masullo Chen ja Pain 2017, 882; Larsson 2018, 607–608; Wolfgang, Blackburn ja McConnell 2020, 443–444). Tämä heijastelee niin sanotun objektiivisen journalismin ihannetta, jossa eksplisiittisesti esitetyt tunteet koetaan journalismin laatua ja puolueettomuutta uhkaavina (Kotisoiva 2019, 15–16). Uutisten verkkokomentointia käsittelevän kirjallisuuden perusteella saattaakin muodostua käsitys, että kaikenlainen tunneilmaisuus olisi epätoivottavaa ja ongelmallista laadukkaan kommentoinnin kannalta. Tunteiden rooliin uutisten verkkokomentoinnissa on kuitenkin keskitytty tutkimuksessa toistaiseksi melko vähän. Lisäksi valtaosa tutkimuksista, joissa uutisten verkkokomentointia käsitellään tunnenäkökulmasta, keskittyy kommenttien asiattomuuteen (engl. *uncivil*, *incivil*). Asiattomuuskehityksen puitteissa verkkokomentoinnin tunteita on tarkasteltu muun muassa kommenttien ja uutisten vaikuttavuuden ja vastaanoton näkökulmasta (esim. Anderson ym. 2014; Gonçalves 2018; Hsuef, Yogeewaran ja Malinen 2015; Naab ym. 2020).

Lisäksi merkittävä osa asiattomuuskeskustelua ovat tutkimukset, jotka käsittelevät kommenttien vaikutusta journalisteihin (esim. Frischlich, Boberg ja Quandt 2017; Metzler 2015; Pöyhtäri, Haara ja Raittila 2013).

Uutisten verkkokommentoinnin tunteita on tarkasteltu lisäksi muun muassa affektiteorioiden näkökulmasta (esim. Nikunen ja Pantti 2018; Papacharissi 2015; Rantasila 2018; Rantasila 2022; Wahl-Jorgensen 2019). Näissä tutkimuksissa tunteet hahmottuvat oleelliseksi osaksi sekä ihmisten laajempia merkitysten muodostamisen tapoja että uutisten kommentointia ja julkista keskustelua (Kotisoiva 2019, 16–17; Wahl-Jorgensen 2019, 45–46, 65). Masullo Chen ja kumppanit (2019, 2) esittävätkin, että asiattomuuden käsitteestä tulisi pyrkiä eroon. Se on heidän mielestään liian epämääräinen, koska tunneilmaisu ja erityisesti negatiivisten tunteiden voimakas ilmaiseminen ovat tietyissä tapauksissa perusteltu osa julkisen mielipiteen ilmaisua (Masullo Chen ym. 2019, 3). Tunteiden moniulotteisuus korostuu erityisesti silloin, kun kommentoitavat uutiset liittyvät kansannousujen ja muun poliittisen liikkehdinnän, luonnonkatastrofien tai suuronnettomuuksien kaltaisiin aiheisiin (Nikunen ja Pantti 2018; Papacharissi 2015; Rantasila 2018; Rantasila 2022). Esimerkiksi sodanvastainen protesti voi sisältää hyvinkin voimakasta tunneilmaisua, mutta liberaalissa demokratiassa sitä pidetään silti legitimiin poliittisen mielipiteen ilmaisuna.

Oman kerroksensa keskusteluun tuo näkökulma tunteista journalististen tekstien autenttisuuden ja aitouden merkitsijöinä (esim. Pantti 2010; Peters 2011; Wahl-Jorgensen 2020). Sama näkökulma toistuu laajemmassa mittakaavassa keskustelussa journalismin niin kutsutusta tunteellistumisesta (Pantti 2010; Peters 2011; Koivunen 2008; Kotisoiva 2019; Wahl-Jorgensen 2020). Tässä artikkelissa ajatellaan verkkokommenttien tunneilmaisua legitimiinä osana julkista keskustelua. Kuten Masullo Chen ja kumppanit (2019), koemme ongelmalliseksi näkökulmat, joissa tunneilmaisu laajasti ymmärrettyinä yhdistetään asiattomaan kommentointiin. Verkkokommentointi voi mielestämme olla tunteikasta olematta asiattonta, ja toisin päin.

## Moderointia valinnan logiikasta hoivan logiikkaan

Uutiskommenttien moderointi voidaan jakaa *esi-* tai *jälkimoderointiin*. Esimoderoinnissa kommenttikenttään lähetettävät viestit käyvät moderointiprosessin läpi ennen kuin ne julkaistaan. Jälkimoderoinnissa kaikki kommentit julkaistaan heti, mutta sopimattomia kommentteja poistetaan jälkikäteen. Molempia moderointitapoja voidaan käyttää sekä ihmisten tekemässä että tietokoneavusteisessa moderoinnissa.

Valtaosa verkkokeskustelujen moderointiratkaisuista perustuu niin kutsutulle *valinnan logiikalle* (Ruckenstein ja Turunen 2020). Valinnan logiikkaa noudattavassa moderoinnissa tavoitteena on poistaa verkkokeskustelun sääntöjä rikkovat tai muuten asiattomat viestit (Gillespie 2018, 5; Gorwa, Binns ja Katzenbach 2020, 3). Anonymiteetin on havaittu lisäävän asiattomien kommenttien määrää (Santana 2014, 26–27; Suh ym. 2018, 194–195). Sen vuoksi kommentoijilta voidaan lisäksi vaatia esimerkiksi kirjautumista joko pseudonyymillä tai omalla nimellä. Tämä hillitsee asiattomuuksia mutta ei täysin poista niitä (Gillespie 2018, 62–63; Santana 2014, 27; Suh ym. 2018, 194–195). Asiattomuuksia on pyritty hillitsemään myös muun muassa estämällä kommentointi aiheissa, joiden tiedetään kirvoittavan run-

saasti asiattomia kommentteja, sekä antamalla lukijoille mahdollisuus ilmiäntää asiattomat viestit (Marchionni 2015, 230). Lisäksi asiattomia kommentteja pyritään tunnistamaan niiden sisällön perusteella. Yksinkertaisimmillaan sisältöön perustuva, valinnan logiikkaa noudattava moderointi perustuu esimerkiksi listoihin, joiden sisältämiä kiellettyjä sanoja tai kuvia joko ihminen tai tietokoneohjelma etsii kommentteista (Gorwa, Binns ja Katzenbach 2020, 4–7). Jos kommentti sisältää yhden tai useampia kiellettyjä sanoja, sen julkaisu joko estetään tai se poistetaan jälkikäteen (emt.). Useimmat suomalaiset mediatilat käyttävät erilaisia yhdistelmiä edellä mainituista keinoista.

Ruckenstein ja Turunen (2020, 1028, 1040) toteavat, että valinnan logiikkaan perustuvat moderointikäytännöt ovat puutteellisia, sillä ne vain siivoavat keskustelujen ongelmalliset elementit pois puuttumatta siihen *miksi* epäasiallisia kommentteja kirjoitetaan. He esittävät, että valinnan logiikkaan perustuvat moderointiratkaisut häivyttävät moderoinnit inhimillisinä toimijoina taustalle ja tuottavat mekanistista näkökulmaa verkkokeskusteluihin (emt., 1040). Ruckenstein ja Turunen (2020, 1027) ehdottavat, että tehokkuuteen ja nopeuteen pyrkivän valinnan logiikan sijaan moderoinnin tulisi noudattaa *hoivan logiikkaa*. Alun perin Molin (2008) esittelemän hoivan logiikan näkökulmaa noudattava moderointi on avoin prosessi, joka edellyttää vuorovaikutusta kaikkien osapuolten välillä (Ruckenstein ja Turunen 2020, 1032). Tällainen moderointi pyrkii aktiivisesti ylläpitämään keskustelua, palkitsemaan hyvistä käytöksistä sekä puuttumaan huonon käytöksen syihin (emt., 1033). Ruckensteinin ja Turusen (2020, 1033–1034) mukaan hoivan logiikkaa seuraava moderointi on tietoisempaa keskustelujen asiayhteyksistä ja lähestyy otteeltaan vuorovaikutteista yhteisön ylläpitoa sen sijaan, että pyrkisi vain siivoamaan ei-toivotun sisällön pois näkyvistä.

Yksi hoivan logiikkaa ja osallistavan journalismin periaatteita noudattava moderointitapa on edellä mainittu journalistien osallistuminen keskusteluun. Esimerkiksi sekä Masullo Chen ja Pain (2017, 882, 888) että Suh ja kumppanit (2018, 197) esittävät, että uutiskomenttien asiallisuuden takaamiseksi ensimmäisen kommentin tulisi joko olla journalistin kirjoittama tai kommenttijärjestelmän tulisi nostaa asiallisin kommentti ensimmäiseksi uutisjutun alle. Kommenttiketjun ensimmäisen viestin on havaittu olevan muiden kommenttien sävyn kannalta olennainen: jos ensimmäinen kommentti on asiallinen ja rakentava, pienenee asiattomien kommenttien todennäköisyys (Gonçalves 2018, 615; Masullo Chen ja Pain 2017, 888; Suh ym. 2018, 197). Lisäksi toimittajien osallistuessa kommenttikeskusteluun sen sävy on pysynyt asiallisempana verrattuna keskusteluihin, joissa toimittaja ei ole läsnä (Graham ja Wright 2015; Ksaizek 2018; Ksaizek, Peer ja Lessard 2016; Masullo Chen ja Pain 2017). Hoivan logiikkaa noudattava moderointi vaatii kuitenkin aikaa, jota usein on uutistoimituksissa hyvin vähän (Reich 2011; Pöyhtäri, Haara ja Raittila 2013). Lisäksi, kuten Gillespie (2018, 74) huomauttaa, kaikkeen julkiseen ja puolijulkiseen keskusteluun liittyy ristiriita keskustelun ylläpidon ja sen sääntelyn välillä. Hoivan logiikka ei poistane tarvetta valinnan logiikkaan perustuvalla moderoinnilla kokonaan.

Ratkaisuksi moderoinnin aikapulaan on Suomessakin tarjottu tietokoneavusteista moderointia, ja se on yleistynyt 2010-luvun puolivälin jälkeen. Esimerkiksi vuonna 2017 *Kaleva*<sup>1</sup> ja tuolloin Aller Median omistuksessa ollut Suomi24-keskustelufoorumi ryhtyivät käyttämään keskusteluidensa moderoinnissa tietokoneavusteista järjestelmää<sup>2</sup>. Suomi24:n ja *Iltalehden* käyttämä ohjelmisto on ostettu Utopia Analyticsiltä<sup>3</sup>, Kalevan puolestaan Vincitiltä. *Helsingin Sanomat* (HS), *Iltalehti* ja *Yleisradio* (YLE) ottivat koneavusteisen moderoinnin käyttöön

hieman myöhemmin. Vuonna 2020 YLE ja HS kuitenkin ilmoittivat luopuvansa koneavusteisesta moderoinnista heikkojen kokemusten takia ja siirtyivät Suomen Tietotoimiston (STT) mediapalveluiden tarjoaman ihmismuotoisen moderointipalvelun asiakkaaksi<sup>4</sup>. Vuoden 2020 aikana myös Keski-suomalainen-konsernin sanomalehdet alkoivat ostaa moderointia STT:ltä<sup>5</sup>.

## Tietokoneavusteisen moderoinnin periaatteet ja sudenkuopat

Valinnan logiikkaa noudattavaa tietokoneavusteista moderointia on hyödynnetty erilaisten estolistojen ja suodattimien muodossa internetin alkuajoista lähtien (Gillespie 2018; Gorwa, Binns ja Katzenbach 2020). 2020-luvulle tultaessa moderointisovellukset hyödyntävät yhä useammin lisäksi erilaisia luonnollisen kielen käsittelyn tai kuvien tunnistamisen järjestelmiä (Gorwa, Binns ja Katzenbach 2020, 3–5). Ne perustuvat tietojenkäsittelytieteellisiin ja tilastollisiin menetelmiin, joiden avulla sisältöä tunnistetaan tai luokitellaan suhteessa muihin moderointiohjelman kohtamiin aineistoihin (emt.). Esimerkiksi sosiaalisen median alustoilla toimii yhtä aikaa useita ohjelmia, jotka arvioivat kuuluuko jaettu kuva tai teksti kiellettyyn kategoriaan (Roberts 2019, 37). Jos ohjelma tunnistaa kuvan tai tekstin kielletyksi, se joko automaattisesti poistetaan tai ohjelma ilmoittaa siitä ihmismoderoijalle (emt.). Sisällön suodattamisen lisäksi koneavusteista moderointia voidaan hyödyntää esimerkiksi tekijänoikeuksien valvontaan tai roskapostin estoon (Grimmelmann 2015; Gorwa, Binns ja Katzenbach 2020, 7–10). Ymmärtääksemme valtaosa suomalaisissa mediataloissa käytössä olevista tietokoneavusteisen moderoinnin sovelluksista noudattaa samoja toimintaperiaatteita kuin Gorwan, Binnsin ja Katzenbachin (2020, 3–5) kuvailemat sosiaalisen median alustoilla käytössä olevat sovellukset. Sovellusten yksityiskohtainen toiminta on kuitenkin liikesalaisuuksien suojaamaa – aiheesta käydään jatkuvaa keskustelua kriittisen algoritmitutkimuksen ja sosiaalisen median tutkimuksen kentillä (ks. esim. Bucher 2018; Gillespie 2018; O’Neil 2016).

Vaikka koneavusteisen moderoinnin sovelluksien toimintaperiaatteet ovat melko yksinkertaisia, ne ovat harvoin ongelmattomia. Usein kritisoidaan esimerkiksi sitä, millaisten materiaalien ja periaatteiden pohjalta aineisto, joilla sovelluksia opetetaan esimerkiksi tunnistamaan ei-toivottua sisältöä, koostetaan (Gorwa, Binns ja Katzenbach 2020, 10; Laaksonen ym. 2020, 11). Esimerkiksi sovelluksille syötettäviä opetusaineistoja koodaavien henkilöiden taustat, asenteet ja kulttuurisidonnaiset arvot voivat vaikuttaa siihen, miten haitalliseksi kommentti luokitellaan (Gerrard ja Thornham 2020, 1269–1270; Laaksonen ym. 2020, 10–11). Tietokoneavusteisen moderoinnin sovellukset eivät myöskään toimi yhtä hyvin kaikilla kielillä (Haapoja, Laaksonen ja Lampinen 2020, 8; Laaksonen ym. 2020, 8). Tämä johtuu yhtäältä siitä, ettei suomeksi ole tarjolla riittävää määrää materiaalia, jolla opettaa sovellusta, ja toisaalta kielten rakenteellisista eroista (Haapoja, Laaksonen ja Lampinen 2020, 8; Laaksonen ym. 2020, 8). Sanalistoihin perustuvaa moderointia on myös helppo kiertää (Chancellor ym. 2016; Gerrard 2018; Grön ja Nelimarkka 2020; Haapoja, Laaksonen ja Lampinen 2020).

Ongelmistaan huolimatta tietokoneavusteinen valinnan logiikkaan nojaava moderoinnista on tullut pysyvä osa keinovalikoimaa, jota suuret kaupallisesti toimivat sosiaalisen



median alustat sekä muut moderointia tarvitsevat tahot käyttävät. Tutkijoiden ja laajemman yleisön suhtautuminen tietokoneavusteiseen moderointiin on kuitenkin kaksijakoista. Teknologian kehitykseen luotetaan, mutta sen heikkouksista ja siihen mahdollisesti sisään rakennetuista epätasa-arvoa tuottavista elementeistä ollaan hyvin tietoisia. Toisin sanoen tekstipohjaisen kielellisen viestinnän kompleksisuus, monimerkityksisyys ja tilannesidonaisuus heikentävät valinnan logiikkaa noudattavan moderoinnin tuloksia. Samat seikat myös motivoivat hoivan logiikan näkökulman huomioimista, riippumatta moderoinnin tekijöistä ja työkaluista. Palaamme tähän huomioon artikkelin lopussa, kun keskustelemme tutkimustulostemme laajemmista vaikutuksista.

## Aineisto ja menetelmät

Artikkelissa analysoitu aineisto koostuu 11 suomalaisen uutismedian verkkokommenttien moderointiin perehtyneen henkilön teemahaastattelusta, joista kymmenen tehtiin vuonna 2019 ja yksi vuonna 2020. Haastatellut valittiin niin, että heistä muodostui suomalaisen uutismedian eri toimijoita vuosina 2019 ja 2020 kattavasti kuvaava joukko. Kymmenen haastatelluista työskenteli haastatteluhetkellä suomalaisissa mediataloissa tehtävissä, joihin verkkokommenttien moderointi liittyi olennaisesti. Lisäksi haastateltiin yhtä suomalaisille mediataloille moderointipalveluita tarjoavan teknologiayrityksen edustajaa, joka on taustaltaan journalisti. Haastatellut henkilöt edustivat viittä keskeisintä suomalaista media-alan konsernia, jotka omistavat yhteensä 97 sanoma- tai kaupunkilehteä, ja joilla on toimintaa yli 30 kunnan ja kaupungin alueella Suomessa ja ulkomailla. Haastatelluista yhdeksän oli miehiä ja kaksi naisia ja he olivat iältään 33–52-vuotiaita. Pseudonymisoimme haastatellut ja erotimme heidät numeroin 1–11. Lisäksi aineistosta nostettuja lainauksia on hieman muokattu luottavuuden vuoksi. Väätäjä ja Rantasila keräsivät haastatteluaineiston ja muut kirjoittajat osallistuivat haastatteluteemojen suunnitteluun sekä osallistujien rekrytointiin.

Tutkimukseen osallistuneita haastateltiin heidän työpäivänsä aikana ja haastattelut kestivät 45–116 minuuttia. Puolistrukturoidun haastattelun kysymykset käsittelivät uutisten verkkokommentteja ja -keskusteluja, niiden moderointia sekä tunteiden merkitystä verkkokeskusteluissa. Haastattelut litteroitiin ja litteroitu materiaali käsiteltiin analyysia varten Atlas.ti-ohjelmistolla. Litteroitu aineisto analysoitiin laadullisen teema-analyysin keinoin ja aineistosta pyrittiin tunnistamaan puhetapoja, joita haastateltavat liittävät moderointiin, tunteisiin ja journalistiseen työhön suhteessa kommentointiin. Nojaamme analyysissamme diskurssianalyttiseen ajatteluun, jossa kielenkäyttöä tarkastellaan toimintana, joka rakentaa sosiaalista todellisuutta (van Dijk 1993). Tarkastelemme, millaisiin asiayhteyksiin uutiskommenttien tunteet ja kommenttien moderointi haastatteluissa kytkeytyvät ja millaisia merkityksiä niille annetaan. Lisäksi analysoimme kommenttien moderoinnin suhdetta journalistiseen työhön.

Analyysiprosessi noudatti aineistolähtöistä hermeneuttista otetta, jossa laadulliset huomioidut muodostuvat ja tarkentuvat sitä mukaa, kun analysoija käsittelee aineistoa toistuvasti (Günther, Hasanen ja Juhila 2020). Ensimmäisessä vaiheessa analysoija luokitteli aineiston laajempia teemoja kuvaavien koodien perusteella. Vaihe tuotti yhteensä 18 koodia. Toi-

sessä vaiheessa kuvailevat koodit jaettiin viiteen yläkategoriaan, joista kuhunkin liitettiin kolmesta viiteen toisiinsa temaattisesti liittyvää koodia. Kolmannessa vaiheessa aineisto luettiin uudelleen läpi useita kertoja lähilukuna, jonka tarkoituksena oli tuottaa mahdollisimman syvällistä tietoa erityisesti moderointiin ja tunteisiin liittyvistä puhetavoista. Tämän jälkeen analyysoijat muodosti kolmannen luokittelun, joka kuvaa haastattelujen moderointiin ja tunteisiin liittyviä laajempia diskursseja. Aineiston analyysistä vastasi pääasiassa Rantasila. Kiskola ja Olsson osallistuivat toiseen ja kolmanteen koodausvaiheeseen.

## Kommentit yleisösuhteen ja eripurakanavana

Seuraavaksi esittelemme analyysimme tärkeimpiä tuloksia. Keskityimme alla neljään aineistosta tunnistettuun teemaan, jotka ovat:

- moderointia koskevat toiveet ja odotukset
- kokemukset moderoinnista
- ihmisten ja koneiden tekemän moderointityön välinen suhde
- moderointi ja tunteet.

Esimoderointi oli haastatteluhetkellä vuonna 2019 käytössä kahdeksassa ja jälkimoderointia käytettiin kahdessa mediatalossa. Tietokoneavusteinen moderointi oli tai oli ollut käytössä puolessa haastateltujen edustamista mediataloista. Lisäksi neljässä mediatalossa ihmisten tekemä moderointi oli haastatteluhetkellä ulkoistettu joko osittain tai kokonaan toimituksen ulkopuolelle.

Valtaosa haastatelluista piti kommentteja oleellisena osana nykyjournalismin yleisösuhtetta ja yleisöille tärkeänä vaikutuskanavana. Lisäksi kaksi haastateltua korosti vilkkaan kommenttikeskustelun tuovan lisää kävijöitä mediatalojen verkkosivuille. Neljä haastateltua kertoi, että kommenttien toivottiin esimerkiksi vievän keskustelua juttujen aiheesta eteenpäin sekä antavan toimittajille uusia näkökulmia ja uutta tietoa. Toisin sanoen lähes kaikki haastatellut kertoivat ensin arvostavansa kommentteja ja kommentoijia edellä mainittujen syiden vuoksi. Samassa yhteydessä kuitenkin tuli ilmi, että kommenttipalstojen todellisuus harvoin vastaa edellä kuvattua ideaalitulannetta.

Kommentoinnin ongelmista puhuttaessa osa haastatelluista nosti trollaamisen ja maalittamisen kaltaisten ilmiöiden lisäksi kommenttien yleisen negatiivisuuden ja heikon laadun (esim. Frischlich, Boberg ja Quandt 2019; Wolfgang 2018). Sen pelättiin karkottavan asialliset mutta harkitsevammat kommentoijat, kuten eräs haastateltu asian alla muotoilee.

*On paljon ihmisiä, jotka seuraavat kommentointia, mutta eivät kirjoita itse, vaikka heillä voisi olla sanottavaa, koska on paljon sitä huonohkoa, ei niin laadukasta kommenttia, niin ihmiset eivät ehkä rohkaistu kirjoittamaan. (Haastateltu 11)*

Toisin sanoen haastatellut vaikuttavat arvostavan vain tietynlaisia kommentteja, jotka vastaavat heidän käsityksiään laadukkaasta tai hyvästä kommentoinnista. Sen sijaan kommentit, jotka eivät vastaa tätä ihannetta, koetaan ongelmallisina ja jopa haitallisina keskustelun laadun kannalta.

Haastateltujen näkemykset toimivasta moderoinnista olivat hyvin yhteneväiset. Heidän näkemyksissään korostuu kommentoija- tai käyttäjälähtöinen ajattelu: moderoinnin tulisi olla nopeaa, tasapuolista ja sen tulisi puuttua kommentoijien ilmaisunvapauteen mahdollisimman vähän. Neljä haastateltua perusteli näkökulmaa sillä, että kommenttipalstan käyttökokemuksen tulisi olla kommentoijille mahdollisimman helppo, joten kommenttien tulisi kulkea ennakkomoderointiprosessin läpi ja tulla julkaistuksi mahdollisimman nopeasti. Samanaikaisesti moderoinnin tulisi ylläpitää kommenttien laatua ja poistaa ei-toivottu sisältö nopeasti, kuitenkin puuttumatta liikaa kommentoijien ilmaisunvapauteen. Kolme haastateltua huomautti, että kommentoijille asetetut vaatimukset sisäänkirjautumisesta ennen kommentointia olivat vähentäneet kommentointia heidän sivustoillaan, vaikka samalla pahimpien asiattomuuksien määrä oli laskenut (ks. Gonçalves 2018; Løvlie, Ihlebæk ja Larsson 2018; Reich 2011).

Haastatelluista neljä korosti kommentoinnin roolia journalismin yleisösuhteen ylläpitäjänä, ja he myös edellyttivät toimittajien osallistuvan keskusteluun kommentoijien kanssa. Seuraava haastatellun huomio tiivistää näkemyksen.

*Mehän tiedetään, että keskustelun laatu paranee, jos toimittajat ovat keskusteluissa tavalla tai toisella läsnä. Se on tavallaan modernin journalismin lähtökohta. (Haastateltu 5)*

Kuten lainaus yllä tiivistää, haastatellut vaikuttivat pitävän toimittajien osallistumista kommenttikeskusteluihin tärkeänä, aikaisempien tutkimustulosten mukaisesti (Wright, Jackson ja Graham 2020). Samassa yhteydessä kävi kuitenkin ilmi, että haastateltujen mielestä keskusteluihin osallistuminen ja muu moderointityö vie paljon aikaa ja rahaa, ja sen koettiin olevan pois toimittajien muusta, usein tärkeämmäksi koetusta työstä (Reich 2011; Wolfgang 2018). Alla näkemystään avaa haastateltu, jonka edustamassa mediatilassa käytettiin haastatteluhetkellä jälkimoderointia, sillä toimiva ennakkomoderointi vaatisi hänen mukaansa huomattavaa rahallista satsausta:

*Kun sen pitäis olla seitsemän päivää viikossa 24 tuntia, tai ainakin merkittävä määrä tunteja vuorokaudessa, niin se on heti kolme neljä viis henkilötyövuotta kiinni siihen hommaan. Ja ennakkomoderointi ei toimi silleen, että "Voitkos sä pääkirjotustoimittaja siinä joutessas niitä kattoo?" Kun se keskustelu kuolee siihen, jos ne viiveet rupee olemaan jotain tunteja. (Haastateltu 8)*

Ratkaisuna ajanpuutteeseen neljässä aineistomme mediatilassa kommenttien moderointi oli ulkoistettu kolmannelle osapuolelle, mutta kuudessa se kuului haastattelujen aikaan toimittajien työtehtäviin ainakin osittain. Kaksi haastatelluista kertoi, että heidän työnantajansa harkitsi kommentoinnin poistamista, koska sen ei koettu tuovan uutisiin lisäarvoa, vaan tuottavan lähinnä eripuraa.

Sen lisäksi, että haastateltavat kokivat moderointityön aikaa vieväksi, osa heistä kuvaili sitä jopa turhauttavaksi. Tämä liittyi haastateltujen huomioissa ennen kaikkea työn suureen määrään sekä aikaisemmin mainittuun kommenttien jatkuvaan, yleiseen negatiivisuuteen. Palaamme tähän huomioon myöhemmin, kun käsittelemme tarkemmin kommenttien tunnesisältöä.

## Koneen kanssa moderoimassa

Tietokoneavusteinen moderointi oli tai oli ollut käytössä puolessa haastateltavien edustamista mediataloista vuonna 2019, ja haastateltujen kokemukset tietokoneavusteisesta moderoinnista olivat monen suuntaisia. Kolme haastateltua ilmaisivat olevansa pääosin tyytyväisiä koneavusteisen moderoinnin toimintaan, ja he kertoivat koneavusteisen ennakkomoderoinnin helpottaneen moderojien työtä ja vähentäneen viivettä kommenttien julkaisussa (vrt. Jhaver ym. 2019). Viiveen pienenemisen myötä näiden mediatalojen kommenttimäärät olivat kasvaneet, joten koneavusteisesta moderoinnista luopuminen myös vaikeutui, kuten eräs haastateltu kertoo alla.

*Kokeiltiin, että oltaisiin otettu viikoksi tekoäly pois käytöstä ja testataan sen jälkeen taas, mutta se oli helpottanut moderojien työtä jo niin paljon ja kommentointimäärät on kasvanu valtavasti, että ei ne moderijat pystyny enää tekemään työtään, kun tuli hirveet jonot.*

(Haastateltu 10)

Toisin sanoen, kun ainakin yhteen edellisen luvun alussa esitetyistä moderoinnin ihanteista (nopeus) vastattiin, kommentoinnin ja sen myötä moderoitavan tekstin määrä lisääntyi. Samalla myös riippuvuus valinnan logiikkaa toteuttavasta koneavusteisesta moderoinnista kasvoi ja moderojien työ vaikutti mukautuvan ohjelmiston määrittämään tahtiin.

Haastatelluista valtaosa suhtautui myös varauksellisesti tietokoneavusteiseen moderointiin ja myönsi, että moderointijärjestelmät vaativat jatkuvaa päivittämistä pysyäkseen kommentoijien luovan kielenkäytön perässä. Myös kommentoijien tietämys moderointijärjestelmien toiminnasta tarkoitti, että moderoinnin oli jatkuvasti oltava valppaana. Huomiot tukevat aikaisemman tutkimuksen tuloksia (mm. Haapoja, Laaksonen ja Lampinen 2020; Tubaro, Casilli ja Coville 2020). Alla olevissa sitaateissa kaksi haastateltua viittaavat samaan ilmiöön.

*Meillä on automaattimoderoinnissa se heikkous, että sinne voi mennä epäasiallinen kommentti läpi jos [...] se on niin älykkäästi tai ovelasti tehty, että tekoäly, vaikka se ottaa huomioon kontekstin, jutun otsikon ja muuta, ei välttämättä pysty tulkitsemaan kaikkea.*

(Haastateltu 2)

*Mä tiedän miten hirvittävän vaikeaa on saada siitä [moderoinnista] aukotonta. Täällä on tietynlainen skeptisyys siihen, et tietty jos onnistuis tekemään sinne täydellisen mallin.*

(Haastateltu 4)

Yhdessä mediatalossa tietokoneavusteinen moderointi oli ollut käytössä, mutta siitä oli tutkimushaastattelun aikaan päätetty toistaiseksi luopua huonojen kokemusten vuoksi. Viisi haastateltua työskenteli mediataloissa, joissa tietokoneavusteista moderointia ei ollut käytössä, mutta he olivat kiinnostuneita tietokoneavusteisesta moderoinnista. He tunnustivat hyvin myös edellä esiin nostetut puutteet. Niissä mediataloissa, joissa tietokoneavusteinen moderointi oli haastatteluhetkellä käytössä, sitä pidettiin puutteineenkin tarpeellisenä ja hyödyllisenä välineenä. Seuraava lainaus tiivistää asenteen, jolla useimmat

haastatellut suhtautuivat tietokoneavusteiseen moderointiin: ”Mut se ei ratkase, se ei oo semmonen autuaaksitekevä juttu, vaan enemmän se avustaa” (Haastateltu 9).

Aikaisempaa tutkimusta mukaillen haastatellut nimesivät tietokoneavusteisten järjestelmien heikkouksiksi kaksi keskeistä asiaa: järjestelmälle syötettävän riittävän suuren koulutusaineiston puutteen ja kommenttien sosiaalisten, kulttuuristen ja poliittisten asiayhteyksien monikerroksisuuden. Erilaisten järjestelmien kouluttamiseksi tarvittavan aineiston määrään liittyvistä ongelmista on keskusteltu niin kotimassa kuin kansainvälisesti sekä tieteellisessä että populaarissa julkisuudessa laajalti viimeisen viiden vuoden aikana, joten voi olla, että haastateltavamme olivat poimineet tämän näkökulman aiheeseen liittyvästä keskustelusta (esim. Haapoja, Laaksonen ja Lampinen 2020). Osalla heistä oli myös henkilökohtaista kokemusta koneavusteisten moderointijärjestelmien kanssa työskentelystä järjestelmien tilaajina, niiden käytön esihenkilöinä tai yhdessä tapauksessa järjestelmän tuottajana. Kommenttien asiayhteyksien monikerroksisuus tuli esille, kun haastatellut kuvailivat, miten moderointijärjestelmiä oli mahdollista kiertää. Usein toistuva esimerkki oli rasistisen solvauksen kirjoittaminen niin, että pari kirjainta vaihdettiin toiseksi, jolloin ihmismoderoiija kyllä tunnisti sanan edelleen solvaukseksi, mutta koneavusteinen järjestelmä ei.

Osa haastatelluista huomautti järjestelmien heikkouksista puhuttaessa, että kommentoinnin kontekstien ymmärtäminen on vaikeaa tietokoneavusteisen moderoinnin järjestelmien lisäksi myös ulkoistetuille ihmismoderoijille, jotka työskentelevät erillään toimituksista ja niiden rutiineista. Kaksi haastatelluista arveli, että tämä johtui sekä moderoijien erilaisista taustoista että yksittäisille moderoijille tulevien kommenttien suuresta määrästä. Kommenttien volyyymi puolestaan teki työstä kaavamaisesta ja saattoi johtaa moderoijien turtumiseen asiattomiin kommentteihin (vrt. Ruckenstein ja Turunen 2020). Erään haastateltavan mukaan kontekstin ymmärtämistä saattaa vaikeuttaa myös moderointijärjestelmän käyttöliittymä: ”Esimerkiksi meillä näkyy vaan otsikko jutusta [moderoitavaksi tulleen kommentin yhteydessä], niin siitä voi olla vaikea päätellä, että onko tämä kommentti nyt asiallinen, kuuluuko se tähän asiaan” (haastateltu 10).

Haastateltujen vastauksista muodostuu käsitys melko pragmaattisesta asenteesta tietokoneavusteista moderointia kohtaan. Koneen tekemän moderoinnin koetaan nopeuttavan ihmismoderoijien työtä, ja moderoinnin nopeutta pidettiin tärkeänä kommentoijien käytökokemuksen ja keskustelun sujuvuuden kannalta. Tällainen ohjelmiston ehdoilla tehtävä moderointityö kuitenkin nähtiin myös kaavamaisena, ja sen voi Ruckensteinia ja Turusta (2020) mukaillen tulkita myös olevan alisteista valinnan binääriselle logiikalle. Ulkoistettua ihmisten tekemää moderointia koskevat huomiot kuitenkin herättävät laajempia kysymyksiä journalistisen työn suhteesta verkkokommenttien moderointiin, valinnan ja hoivan logiikkaan sekä moderoinnin ja kommentoinnin käyttöliittymiin. Palaamme näihin kysymyksiin artikkelimme loppupuolella.

## **Moderointi ja kommenttien tunteet**

Haastatellut hahmottivat tunteiden ilmaisun ja tunteiden viriämisen merkittäväksi osaksi uutisten verkkokommentointia. Osa haastatelluista ajatteli, että räiskyvä tunneilmaisu on

itse asiassa olennainen osa uutisten verkkokomentointia, jota ilman keskustelu ei houkuttelisi osallistumaan yhtä innokkaasti, kuten sitaatti alla osoittaa.

*Tietyllä tavalla [tunteet] varmaan ruokkii ja värittää ja piristää sitä keskustelua. Jos siellä ainoastaan argumentoitaisiin hyvin asialähtöisesti niin [...] uskon, että se voisi olla joidenkin silmissä vähän puisevampaa se keskustelu. (Haastateltu 2)*

Tästä huolimatta useat haastateltavat nostivat sekä suoraan että epäsuorasti negatiivisten tunteiden voimakkaan ilmaisuuden yhdeksi syyksi moderoinnin tarpeelle.

*Joo, kaikki tunteet on ok. Mut tietysti se ois aina ihanteellisempaa et me ihmiset tunnistettais paremmin omii tunteita. Se kyllä kehittäis sitä meidän ajattelua myös niissä tilanteissa ja se ehkä kehittäis, se nostais sen keskustelun tasoa. (Haastateltu 1)*

Toisin sanoen kaikki tunneilmaisu ei ollut haastateltujemme mielestä asiattonta. He vetivät asiattomuuden rajan useimmiten siihen, mihin tai kehen kommentit kohdistuivat. Esimerkiksi politiikan asiakysymyksiin liittyvä tyytymättömyyden ilmaisu oli haastatelluista hyväksyttävää, mutta henkilöihin kohdistuvat negatiiviset tunteen purkaukset koettiin ongelmallisiksi. Lisäksi kuusi haastateltua piti yhtä ongelmallisena jankkaamista tai vakio-kommenttien välistä negatiivista nokittelua, joka ei välttämättä liittynyt mitenkään jutun aiheeseen. Tämä nähtiin myös moderoinnin näkökulmasta hankalana: jankkaavat kommentit eivät välttämättä sisältönsä puolesta olleet asiattomia, mutta ne eivät myöskään tuoneet keskusteluun uutta näkökulmaa ja veivät tilaa muilta (potentiaalisilta) keskustelijoilta. Toisin sanoen haastateltumme vaikuttavat jakavan Masullo Chenin ja kumppaneiden (2019) ja Mouffen (1999) näkemyksen, jonka mukaan negatiiviset tunteet ja niiden ilmaisu kuuluvat julkiseen keskusteluun. Eri asia on, mihin tai kehen nuo tunteet kohdistuvat ja miten ne esitetään.

Tähän nähden ristiriitaista oli, että useampi haastateltu piti kommenttien negatiivisuutta ongelmana. Haastatellut eivät kuitenkaan tarkentaneet, mitä negatiivisuus konkreettisesti tarkoitti, vaan se jäi heidän puheessaan yleisen tason havainnoksi. Negatiivisuus kuitenkin mainittiin usein edellä mainitun jankkaamisen yhteydessä, joten tulkitsemme sen tarkoittavan ennen kaikkea paikallaan junnavaa keskustelua, joka ei enää liity kommentoinnin alkuperäiseen aiheeseen.

Positiivisiksi miellettyjen tunteiden ilmaisu nousi esiin erityisesti urheilu-uutisten tai muiden, yleensä laajaa kansallista huomiota keränneiden mediatapahtumien yhteydessä. Lisäksi osa haastatelluista nosti esiin paikallisiin uutisiin liittyviä niin sanottuja hyvän mielen juttuja, joihin liittyi korostuneesti positiivisten tunteiden ilmaisu. Näissäkin tapauksissa haastateltuja pohditutti, kuinka paljon sisällöltään samanlaisia kannustusviestejä esimerkiksi urheilumenestyksestä kertovan uutisen kommentteissa kannatti julkaista ilman, että siitä tuli toisteista. Moderoinnin näkökulmasta kommenttien tunneilmaisuuden kontekstisidonnaisuus ja vaihtelevuus vaikuttaa haastavalta, sillä tässä tapauksessa moderointi ei voi perustua yksittäisten sanojen suodattamiseen vaan vaatii laajempien asiayhteyksien hahmottamista.

## Moderoinnin, teknologian ja tunteiden rajapinnan neljä ristiriitaa

Kaiken kaikkiaan haastatteluista avautuu monitahoinen kuva moderointityöhön suomalaisissa uutistoimituksissa 2020-luvun kynnyksellä. Kuva kuitenkin sisältää useita ristiriitoja. Ensimmäinen ristiriita liittyy moderoinnin ja muun toimituksissa tehtävän journalistisen työn suhteeseen. Aineiston perusteella moderointi hahmotetaan erillisenä tehtävänä suhteessa journalistiseen työhön, vaikka kaikki haastatellut pitivät uutiskommentteja journalismiin vakiintuneena yleisövuorovaikutuksen muotona. Erillisyyttä korosti moderoinnin ulkoistaminen, sillä ulkopuoliset moderoijat eivät haastattelujen perusteella pystyneet kommunikoimaan suoraan toimituksen kanssa. Tätä näkökulmaa vahvistaa vuoden 2020 aikana useissa suomalaisissa mediataloissa tehty päätös kommenttien moderoinnin ulkoistamisesta STT:n moderointipalveluille.

Toinen ristiriita on, että haastatellut odottivat uutiskommenteilta asiakeskeistä ja järjeleivää otetta, mutta kommenttien haluttiin samalla olevan viihdyttäviä ja tyyliltään epämuodollisia. Liian muodollisuuden pelättiin tyrehtyttävän keskustelun ja karkottavan kommentoijat aiheen parista. Lisäksi valtaosa haastatteluista piti negatiivisia tunteen ilmaisuja, kuten pettymystä politiikkaan, legitiiminä osana julkista keskustelua. Samaan aikaan kuitenkin huomattava osa haastatteluista nimesi hieman epämääräisesti kommenttien negatiivisuuden ja huonon ilmapiirin tekijäksi, joka lisäsi kommenttipalstojen moderoinnin tarvetta ja teki moderointityöstä epämiellyttävää. Tämä ristiriita toistaa aiemman tutkimuksen esittämiä väitteitä tunteiden paikasta journalistisessa julkisuudessa (Kotisoiva 2019; Wahl-Jorgensen 2019).

Kolmas ristiriita näkyy suhteessa tietokoneavusteiseen ja ihmisten tekemään moderointityöhön. Tietokoneavusteisen ja toimituksista ulkoistetun moderoinnin koettiin nopeuttaneen moderointiprosesseja, mutta molemmissa nähtiin ongelmia erityisesti keskustelukontekstin ymmärtämisen puutteen ja kommentoijien käyttämien kiertoilmausten takia. Haastateltavien varauksellinen suhtautuminen tietokoneavusteiseen moderointiin heijastelee muun muassa Ruckensteinin ja Turusen (2020) tuloksia. Heidän haastattelemansa keskustelufoorumien moderoijat kokivat tietokoneavusteisen järjestelmän nopeuttaneen heidän työtään, mutta järjestelmä teki työn myös entistä mekaanisemmaksi ja vei sitä kauemmas hoivan logiikan periaatteista (emt.). Haastateltujemme huomiot moderoinnin kaavamaisuudesta ja tehokkuuden tavoittelusta viittaavat samansuuntaiseen kehitykseen. Tietokoneavusteinen moderointi pyrkii tehostamaan moderoijien työtä, mutta nopeutuneen moderoinnin seurauksena moderoijan käsittelemien kommenttien määrä kasvaa, ja moderoijan työ- ja ajattelutavat saattavat alkaa noudatella entistä enemmän mekanistista valinnan logiikkaa. Tällaisesta tehokkuutta ja binäärisiä valintoja suosivasta moderoinnin tavasta voi olla vaikea siirtyä kohti keskustelelevampaa, hoivan logiikkaa noudattelevaan moderointia ilman satsauksia esimerkiksi moderoijien määrään. Lisäksi tietokoneavusteista moderointia koskevissa havainnoissa korostuu kaksijakoinen suhtautuminen uuteen teknologiaan ja sen tuomiin mahdollisuuksiin. Uudelta teknologialta odotettiin paljon, mutta käytännön ongelmat johtivat pettymyksiin, kun odotuksia ei lunastettu. Tämä on tyypillistä uuden teknologian käyttöönotossa, ja osoittaa, että tietokoneavusteisen moderoinnin järjestelmät hakevat edelleen paikkaansa suomalaisissa mediataloissa.

Neljäs moderointiin liittyvä ristiriita näyttäytyy siinä, miten toimittajien keskusteluun osallistumiseen asennoidutaan. Kommentteja pidetään olennaisena osana modernia journalismia. Toimittajien osallistumista on esitetty moderointiratkaisuksi, joka noudattaisi hoivan logiikkaa, sillä keskusteluun osallistuessaan toimittaja myös osallistuu kommentoijista muodostuvan yhteisön rakentamiseen (vrt. Ahva, Hautakangas ja Haara 2017; Graham ja Wright 2015). Aineistostamme kuitenkin käy ilmi, että ainakin Suomessa harvalla toimittajalla vaikuttaa olevan aikaa tai halua osallistua esimerkiksi omia juttujaan koskevaan kommentointiin ainakaan mediatalon kommenttipalstoilla (vrt. Heinonen 2011; Pöyhtäri, Haara ja Raittila 2013). Tämä ei kerro kuitenkaan siitä, miten toimittajat viestivät yleisönsä kanssa muualla verkossa.

Uutiskommentteja käsittelevässä kirjallisuudessa toistuu näkemys kommenteista osana journalistisen tiedonhankinnan prosessia (esim. Park ym. 2016). Kommenttien nähdään tarjoavan toimittajille uusia näkökulmia tai uutta tietoa työn alla olevaan juttuun (emt.). Kommenttipalstojen tärkeyttä perustellaan usein myös lukijapalautteen väylänä (esim. Wright, Jackson ja Graham 2020). Kun kommenttien moderointi ulkoistetaan kolmannelle osapuolelle, yhteys lukijoiden ja toimitusten välillä kuitenkin katkeaa. Aineistomme perusteella vaikuttaa siltä, että kirjallisuudessa kuvatut tapaukset, joissa kommenteista olisi saatu uusi aihe esimerkiksi jatkojuttuun, ovat suomalaisissa uutistoimituksissa melko harvinaisia ja yksittäisten toimittajien aktiivisuuden varassa. Tilannetta pidetään ideaalina, mutta siihen harvoin päästään. Tämä heijastaa aikaisemman tutkimuksen havaintoja siitä, että osalle mediataloista verkkokeskusteluihin osallistuminen ja niiden moderointi on olennainen osa journalismin perustyötä, osa taas näkee sen journalistisesta työstä erilliseksi tehtäväksi (Masullo Chen ja Pain 2017; Wolfgang, McConnell ja Blackburn 2020; vrt. Heinonen 2011; Pöyhtäri, Haara ja Raittila 2013). Toisin sanoen tietokoneavusteinen moderointi ja ulkoistettu moderointityö vapauttavat uutistoimitusten resursseja journalistiseen perustyöhön, mutta samalla kommenteissa käytävät keskustelut jäävät toimituksille etäisiksi, vaikka niiden olemassaoloa perustellaan toimituksen ja lukijoiden välisellä vuorovaikutuksella.

## Tietokoneavusteisella moderoinnilla kohti hoivan logiikkaa?

Haastateltujemme vastaukset vaikuttavat vahvistavan aikaisempia tuloksia, joiden mukaan ihmisten luova kielenkäyttö, mahdollisten merkitysten ja asiayhteyksien monikerroksisuus ja tunneilmaisun moninaisuus tekevät uutiskommenteista koneavusteisen moderoinnin kannalta haastavia. Lisäksi uutiskommenttien reaktiivisuuden ja räiskyvän tunneilmaisun ero suhteessa järkeilevän ja punnitun julkisen keskustelun ihanteeseen herättää useampia kysymyksiä sekä moderoinnin, lukijoille annettavan kommentoinnin ohjeistuksen että kommentoinnin käyttöliittymien kannalta.

Aikaisempaan tutkimukseen nojaten (mm. Gillespie 2018; Ruckenstein ja Turunen 2020) esitämme, että voimakkaimpia tunneilmaisultaan negatiivisia ja henkilöön kohdistuvia kommentteja voisi ennaltaehkäistä hoivan logiikkaa mukailevien moderointikäytäntöjen avulla ilman merkittäviä teknologisia muutoksia moderointijärjestelmiin. Hoivan logiikkaan noudattavien moderointikäytäntöjen, kuten toimittajien aktiivinen osallistuminen kommentti-keskusteluun, laajempi käyttöönotto edellyttää kuitenkin näkökulman muutosta suhteessa moderointiin osana journalistista julkisuutta. Esitämme, että moderointityö tulisi nähdä



osana journalistista työtä ja yleisösuhteen ylläpitoa sen sijaan, että sen ajateltaisiin vievän aikaa ja tilaa journalismin perustehtäviltä. Toisin sanoen, moderointityö tulisi nähdä kokonaisuudessaan journalistisen julkisuuden elementtinä, jota on syytä vaalia ja johon on syytä sijoittaa sekä aikaa että rahaa. Tämä kuitenkin edellyttäisi uutistoimituksilta melko perusteellista oman roolinsa uudelleenarviointia osana 2020-luvun hybridiä mediajulkisuutta.

Moderointityö on usein myös kuormittavaa (Ruckenstein ja Turunen 2020; Roberts 2019). Koneavusteinen ja valinnan logiikkaan pohjaava moderointi koetaan tehokkaana tapana moderoida suuria kommenttimääriä nopeasti, ja se vaikuttaa vakiinnuttaneen paikkaansa osana suomalaisten mediatalojen moderointikäytäntöjä. Haastateltujemme kokemukset antavat ymmärtää, että koneavusteinen moderointi ei toistaiseksi yksin riitä, jos tavoitteena on ylläpitää laadukasta julkista keskustelua. Siksi esitämme toisaalla (Kiskola ym. 2021, Rantasila ym. 2022) mahdollisia suunnitteluperiaatteita, jolla hoivan logiikan voisi tulevaisuudessa tuoda osaksi moderointityötä tavalla, joka ei kuormita ihmismoderoijia. Tällaisia keinoja ovat muun muassa moderointiohjelma, joka tunnistaa kommenttien sävyn sentimenttianalyysin tai luonnollisen kielen tunnistuksen keinoin ja neuvoo kommentoijaa muokkaamaan tekstiään tarvittaessa. Erilaisia suunnitteluvaihtoehtoja on myös helppo muokata moderoitavan keskustelualustan toivotun ilmapiirin mukaan. Käyttöliittymäratkaisujen avulla alustoille voidaan tuoda esimerkiksi leikkimielisiä elementtejä tai muokata niitä asiallisemmiksi (Kiskola ym. 2021; Rantasila ym. 2022). Masullo Cheniä ja muita (2019) mukailen haluamme huomauttaa, että hoivan logiikkaa noudattavaa moderointia ei pidä tulkita yritykseksi poistaa erimielistä puhetta verkosta. Negatiivisten tunteiden ilmaiset kuuluvat julkiseen keskusteluun eivätkä siis itsessään ole ongelmallisia. Julkisen keskustelun ilmapiiri-ongelmat ovat ratkaistavissa, jos niihin halutaan tarttua vakavasti.

## Lopuksi

Uutisten kommentointi suomalaisten mediatalojen verkkosivuilla tuskin tulee vähenemään tulevaisuudessa. Siksi on toivottavaa, että moderointiin suhtauduttaisiin nykyistä vahvemmin osana julkisen keskustelun tuottamista ja ylläpitoa journalismissa. Kuten aiemmin toteamme, tämä kuitenkin edellyttäisi, että mediatalat muuttaisivat suhtautumistaan uutisten verkkokomentointiin ja niiden moderointiin. Haastatteluaineistomme perusteella se on vaikeaa, sillä uutisten verkkokommentteihin ja niiden moderointiin näyttää kiteytyvän lukuisia 2020-luvun hybridin mediajärjestelmän sisäisiä jännitteitä, alkaen sanan- ja ilmaisunvapauden ja sen rajoittamiseen liittyvistä eettisistä kysymyksistä ja päättyen siihen, kuinka moderointityötä arvostetaan toimituksissa.

Sanan- ja ilmaisunvapauden vaalimisen ja uutiskomenttien moderoinnin välinen jännite, sekä kysymykset kommenttien tunteikkuudesta näyttäytyvät tästä näkökulmasta haastatteluaineistossamme erityisen kiinnostavina. Journalistien näkökulmasta on yhtäältä ymmärrettävää, että he asemoituvat epäsuorasti sangen laajan sananvapauden kannattajiksi, eivätkä siksi kannata voimakasta moderointia. Toisaalta he kuitenkin esittävät monenlaisia toiveita kommenttien sisällön suhteen, mutta vaikuttavat suorastaan voimattomilta pohiessaan, kuinka tarttua esimerkiksi toistuvasti negatiivisia kommentteja heittelevien henkilöiden toimintaan.

Esitämme sekä omaan aikaisempaan tutkimukseemme (Kiskola ym. 2021; Rantasila ym. 2022) että Ruckensteiniin ja Turuseen (2020) nojaten, että hoivan logiikkaa mukailevien moderointikäytäntöjen huomioon ottaminen on yksi mahdollinen keino saavuttaa tasa-paino edellä mainittujen jännitteiden välillä. Yllä kuvaamamme kommenttien tunteita ja sävyjä tunnistava tietokoneavusteinen moderointiohjelma tarjoaisi mahdollisuuden puuttua tekstin asiattomuuteen tai negatiivisuuteen ennaltaehkäisevästi, ja antaisi kommentoijalle mahdollisuuden vaikuttaa kommenttinsa julkaisukelpoisuuteen (Rantasila ym. 2022). Tätä voisi vielä vahvistaa aikaisemmissa tutkimuksissa (esim. Graham ja Wright 2015) hyväksi havaitulla tavalla antamalla ihmismoderoijalle mahdollisuus osallistua käynnissä olevaan keskusteluun ja tarvittaessa puuttua siihen, mikäli esimerkiksi keskustelu muuttuu jankkaukseksi. Teknologia ja keinot verkkokommenttien vaalimiseen ovat toisin sanoen jo olemassa, vaaditaan vain rohkeutta ja tahtoa tarttua niihin.

### Viitteet

- 1 Ilmari Ahonen. 2017. "Tekoälykstä moderointia." Vicnit blogi 13.12.2017. Luettu 21.4.2021. <https://www.vincit.fi/fi/tekoalykasta-moderointia/>.  
Noora Autio 2018. "Robotti siivoaa Kalevan keskustelupalstaa – toimittajille vapautuu aikaa moderoinnista juttujen tekoon". Suomen Lehdistö 7.2.2018. Luettu 21.4.2021. <https://suomenlehdisto.fi/robotti-siivoaa-kalevan-keskustelupalstaa-toimittajille-vapautuu-moderoinnista-aikaa-juttujen-tekoon/>.
- 2 "Uusi aluevaltaus – tekoäly keskustelumoderaattorina: "Saadaan vähitellen vastaamaan lähes sataprosenttisesti ihmismoderointia" Markkinointi & Mainonta 14.2.2017 Luettu 21.4.2021. <https://www.marmai.fi/uutiset/uusi-aluevaltaus-tekoaly-keskustelumoderaattorina-saadaan-vahitellen-vastamaan-lahes-sataprosenttisesti-ihmismoderointia/374493ed-914b-3170-963a-f2e62fc0f6ee>.  
Kaisa Heinämäki 2017. "Kun tekoäly oppi moderoimaan." Dingle blogi 25.4.2017. Luettu 21.4.2021. <http://blog.dingle.fi/kun-tekoaly-oppi-moderoimaan>.
- 3 Utopia Analytics. "Case Iltalehti" Luettu 21.4.2021. <https://utopiaanalytics.com/case/case-iltalehti/>.
- 4 Esa Mäkinen 2019. "Kolmen projektin kokemuksella tekoälyhankkeiden arki on kaukana uhka- ja juhlapuheista." Helsingin Sanomat 19.1.2019. Luettu 21.4.2021. <https://www.hs.fi/mielipide/art-2000005970691.html>.  
Esa Mäkinen 2020. "HS:n artikkelikommentointi siirtyi STT:n hoidettavaksi." Helsingin Sanomat 15.1.2020. Luettu 21.4.2021. <https://www.hs.fi/kotimaa/art-2000006373199.html>.  
Tommi Peltoniemi. 2020. "STT:n moderointi laajenee – Yle uudeksi asiakkaaksi, moderointiperiaatteet julkisiksi." STT-Lehtikuva 20.1.2020. Luettu 27.11.2020. <https://stt.fi/stt-moderointi-yle-hs/>.  
Ville Seuri 2020. "Tekoäly sai potkut Yleltä, mikä on konejournalismin tulevaisuus? "Tekoäly ei pysynyt ihmisen luovuuden perässä." Yle Uutiset 25.1.2020. Luettu 21.4.2021. <https://yle.fi/uutiset/3-11171589>.
- 5 Tommi Peltoniemi. 2020. "STT alkaa moderoida Keski-suomalainen-konsernin artikkelikommentteja – yhteistyössä yli 60 lehteä." STT-Lehtikuva 4.2.2020. Luettu 5.2.2021. <https://stt.fi/stt-alkaa-moderoida-keski-suomalainen-konsernin-artikkelikommentteja/>.

### Kirjallisuus

- Ahva, Laura, Mikko Hautakangas ja Paula Haara. 2017. *Sovittelujournalismin käsikirja*. Tampere: Tampereen yliopisto. Luettu 25.5.2023. <http://urn.fi/URN:ISBN:978-952-03-0618-2>.
- Anderson, Ashley A., Dominique Brossard, Dietram A. Scheufele, Michael A. Xenos ja Peter Lawdig. 2014. "The 'Nasty Effect': Online Incivility and Risk Perceptions of Emerging Technologies." *Journal of Computer-Mediated Communication* 19: 373–387. <https://doi.org/10.1111/jcc4.12009>.
- Boberg, Svenja, Tim Schatto-Eckrodt, Lena Frischlich ja Thorsten Quandt. 2018. "The Moral Gatekeeper? Moderation and Deletion of User-Generated Content in a Leading News Forum." *Media and Communication* 6 (4): 58–69. <https://doi.org/10.17645/mac.v6i4.1493>.
- Bucher, Taina. 2018. *IF...THEN: Algorithmic power and politics*. Oxford University Press.

- Chancellor, Stevie, Jessica Pater, Trustin Clear, Eric Gilbert ja Munmun De Choudhury. 2016. "#thyghgapp: Instagram Content Moderation and Lexical Variation in Pro-Eating Disorder Communities." *CSCW '16: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*: 1201–1213. <https://doi.org/10.1145/2818048.2819963>.
- Frischlich, Lena, Svenja Boberg ja Thorsten Quandt. 2019. "Comment Sections as Targets of Dark Participation? Journalists' Evaluation and Moderation of Deviant User Comments." *Journalism Studies*, 20 (14): 2014–2033. <https://doi.org/10.1080/1461670X.2018.1556320>.
- Gerrard, Ysabel. 2018. "Beyond the hashtag: circumventing content moderation on social media." *New Media & Society* 20 (12): 4492–4511. <https://doi.org/10.1177/1461444818776611>.
- Gerrard, Ysabel ja Helen Thornham. 2020. "Content moderation: Social media's sexist assemblages." *New Media & Society* 22 (7): 1266–1286. <https://doi.org/10.1177/1461444820912540>.
- Gillespie, Tarleton. 2018. *Custodians of the Internet. Platforms, Content moderation and the hidden decisions that shape social media*. New Haven & London: Yale University Press.
- Graham, Todd ja Wright, Scott. 2015. "A Tale of Two Stories from "Below the Line": Comment Fields at the Guardian." *The International Journal of Press/Politics* 20 (3): 317–338. <https://doi.org/10.1177/1940161215581926>.
- Grimmelmann, James. 2015. "The virtues of moderation." *Yale Journal of Law & Technology* 17 (42): 42–109. Luettu 25.5.2023. <https://yjolt.org/virtues-moderation>
- Gonçalves, João. 2018. "Aggression in news comments: how context and article topic shape user-generated content." *Journal of Applied Communication Research* 46 (5): 604–620. <https://doi.org/10.1080/00909882.2018.1529419>.
- Gorwa, Robert, Reuben Binns ja Christian Katzenbach. 2020. "Algorithmic content moderation: Technical and political challenges in the automation of platform governance." *Big Data & Society* 7 (1): 1–15. <https://doi.org/10.1177/2053951719897945>.
- Grön, Kirsikka ja Nelimarkka, Matti. 2020. "Party Politics, Values and the Design of Social Media Services: Implications of political elites' values and ideologies to mitigating of political polarisation through design." *Proceedings of the ACM on Human-Computer Interaction* 4 (2): 1–29. <https://doi.org/10.1145/3415175>.
- Günther, Kirsi, Kirsi Hasanen ja Kirsi Juhila. 2020. "Analyysi ja tulkinta." Teoksessa *Laadullisen tutkimuksen verkkokäsikirja*, toimittanut Jaana Vuori. Tampere: Yhteiskuntatieteellinen tietoaarkisto. Luettu: 26.10.2021. <https://www.fsd.tuni.fi/palvelut/menetelmaopetus>.
- Haapoja, Jesse, Salla-Maaria Laaksonen ja Airi Lampinen. 2020. "Gaming Algorithmic Hate-Speech Detection: Stakes, Parties, and Moves." *Social Media + Society* 6 (2): 1–10. <https://doi.org/10.1177/2056305120924778>.
- Haara, Paula. 2012. "Poliittinen maahanmuuttokeskustelu Helsingin Sanomien verkkokeskusteluissa." Teoksessa *Maahanmuutto, media ja eduskuntavaalit*, toimittanut Mari Maasilta, 52–86. Tampere: Tampere University Press.
- Harju, Auli. 2018. "Suomiz4-keskustelut kohtaamisten ja törmäysten tilana." *Media & viestintä* 41 (1): 51–74. <https://doi.org/10.23983/mv.69952>.
- Heinonen, Ari, Maarit Mäkinen, Seija Ridell, Ari Martikainen, Mika Halttu ja Esa Sirkkunen. 2000. *Verkkotorilla. Internet kansalaisviestinnän ja paikallisen julkisuuden tilana. Paikallisuus verkkomediassa -projektin loppuraportti*. Tampereen yliopisto. Tiedotusopin laitos. Journalismin tutkimusyksikkö. Julkaisuja C 32. Luettu 26.5.2023. <https://urn.fi/urn:isbn:951-44-4998-3>.
- Heinonen, Ari. 2011. "The journalist's relationship with users: New dimensions to conventional roles." Teoksessa *Participatory journalism: Guarding open gates at online newspapers*, toimittaneet Jane B. Singer, David Domingo, Ari Heinonen, Alfred Hermida, Steve Paulussen, Thorsten Quandt, Zvi Reich ja Marina Vunjavic. 34–56. London: Wiley-Blackwell.
- Hsuef, Mark, Kumar Yogeewaran, ja Sanna Malinen. 2015. "'Leave Your Comment Below': Can Biased Online Comments Influence our Own Prejudicial Attitudes and Behaviors?" *Human Communication Research* 41: 557–576. <https://doi.org/10.1111/hcre.12059>.
- Jhaver, Shagun, Iris Birman, Eric Gilbert ja Amy Bruckman. 2019. "Human-Machine Collaboration for Content Regulation: The Case of Reddit Automoderator." *ACM Transactions on Computer-Human Interaction* 26 (5): 1–35. <https://doi.org/10.1145/3338243>.
- Juntunen, Jussi. 2008. *Verkkoyhteisöjen moderointi. Keskustelufoorumit Näkökulma ja Muusikoiden.net*. Tietojenkäsittelytieteen Pro Gradu -työ, Oulun yliopisto. Luettu 10.2.2021. <http://docplayer.fi/4296393-Verkkoyhteisöjen-moderointi-keskustelufoorumit-nakokulma-ja-muusikoiden-net.html>.
- Kangaspunta, Veera. 2021. *Verkkouutisten kommentit julkisen osallistumisen muotona ja julkisen rajapinnoilla*. Tampereen yliopiston väitöskirjat 362. Vantaa: PunaMusta Yliopistopaino. Luettu: 25.5.2023. <http://urn.fi/URN:ISBN:978-952-03-1818-5>.

- Kiskola, Joel, Thomas Olsson, Heli Väättäjä, Aleksi H. Syrjämäki, Anna Rantasila, Poika Isokoski, Mirja Ilves ja Veikko Surakka. 2021. "Applying Critical Voice in Design of User Interfaces for Supporting Self-Reflection and Emotion Regulation in Online News Commenting." *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-13. <https://doi.org/10.1145/3411764.3445783>.
- Koivunen, Anu. 2008. "Affektin paluu? Tunneongelma suomalaisessa mediatutkimuksessa." *Tiedotustutkimus* 31 (3): 5-24. <https://doi.org/10.23983/mv.63018>.
- Kotisoiva, Johana. 2019. "The elephant in the newsroom: Current research on journalism and emotion." *Sociology Compass*. 13 (5): 1-11. <https://doi.org/10.1111/soc4.12677>.
- Ksaizek, Thomas B. 2018. "Commenting on the News." *Journalism Studies* 19 (5): 650-673. <https://doi.org/10.1080/1461670X.2016.1209977>.
- Ksaizek, Thomas B., Limor Peer ja Kevin Lessard. 2016. "User engagement with online news: Conceptualizing interactivity and exploring the relationship between online news videos and user comments." *New Media & Society* 18 (3): 502-520. <https://doi.org/10.1177/1461444814545073>.
- Laaksonen, Salla-Maaria, Jesse Haapoja, Teemu Kinnunen, Matti Nelimarkka ja Reeta Pöyhtäri. 2020. "The Datafication of Hate: Expectations and Challenges in Automated Hate Speech Monitoring." *Frontiers in Big Data* 3 (3): 1-16. <https://doi.org/10.3389/fgdata.2020.00003>.
- Larsson, Anders Olof. 2018. "Assessing 'The Regulars'—and Beyond." *Journalism Practice* 12 (5): 605-623. <https://doi.org/10.1080/17512786.2017.1338149>.
- Løvlie, Anders Sundnes, Karoline Andrea Ihlebæk ja Anders Olof Larsson. 2018. "User Experiences with Editorial Control in Online Newspaper Comment Fields." *Journalism Practice* 12 (3): 362-381. <https://doi.org/10.1080/17512786.2017.1293490>.
- Marchionni, Doreen. 2015. "Online story commenting. An experimental test of conversational journalism and trust." *Journalism Practice* 9 (2): 230-249. <https://doi.org/10.1080/17512786.2014.938943>.
- Masullo Chen, Gina ja Pain, Paromita. 2017. "Normalizing Online Comments." *Journalism Practice*, 11:7, 876-892. <https://doi.org/10.1080/17512786.2016.1205954>.
- Masullo Chen, Gina, Ashley Muddiman, Tamar Wilner, Eli Pariser ja Natalie Jomini Stroud. 2019. "We Should Not Get Rid of Incivility Online." *Social Media + Society*, 5 (3): 1-5. <https://doi.org/10.1177/2056305119862641>.
- Meltzer, Kimberly. 2015. "Journalistic Concern about Uncivil Political Talk in Digital News Media: Responsibility, Credibility, and Academic Influence." *The International Journal of Press/Politics* 20 (1): 85-107. <https://doi.org/10.1177/1940161214558748>.
- Mol, Annemarie. 2008. *The Logic of Care: Health and the Problem of Patient Choice*. London: Routledge.
- Mouffe, Chantal. 1999. "Deliberative Democracy or Antagonistic Pluralism?" *Social Research* 66 (3): 745-759.
- Naab, Teresa K., Dominique Heinbach, Marc Ziegele ja Marie-Theres Grasberger. 2020. "Comments and Credibility: How Critical User Comments Decrease Perceived News Article Credibility." *Journalism Studies* 21 (6): 783-801. <https://doi.org/10.1080/1461670X.2020.1724181>.
- Nikunen, Kaarina ja Pantti, Mervi. 2018. "Tapaus Ku Klux Klan: affektiivinen julkisuus, moraaliset tunteet ja tahmaiset kuvat." Teoksessa *Pakolaisuus, tunteet ja media*, toimittaneet Mari Maasilta ja Mervi Pantti, 71-91. Tampere: Vastapaino
- O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Random House.
- Pantti, Mervi. 2010. "The Value of Emotion: An Examination of Television Journalists' Notions on Emotionality." *European Journal of Communication* 25 (2): 168-181. <https://doi.org/10.1177/0267323110363653>.
- Papacharissi, Zizi. 2010. *A Private Sphere. Democracy in a Digital Age*. Cambridge: Polity.
- Papacharissi, Zizi. 2015. *Affective Publics: Sentiment, Technology, and Politics*. Oxford: Oxford University Press
- Park, Deokgun, Simranjit Sachar, Nicholas Diakopoulos ja Niklas Elmqvist. 2016. "Supporting Comment Moderators in Identifying High Quality Online News Comments." *CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, May 2016, 1114-1125. <https://doi.org/10.1145/2858036.2858389>.
- Peters, C. 2011. "Emotion Aside or Emotional Side? Crafting an 'Experience of Involvement' in the News." *Journalism: Theory, Practice & Criticism* 12 (3): 297-316. <https://doi.org/10.1177/1464884910388224>.
- Pietilä, Veikko ja Ridell, Seija. 1998. "Julkea ehdotus." *Tiedotustutkimus* 21 (1): 95. Luettu 25.5.2023. <https://journal.fi/mediaviestinta/article/view/61350>.
- Pöyhtäri, Reeta, Paula Haara ja Pentti Raittila. 2013. *Vihapuhe sananvapautta kaventamassa*. Tampere: Tampere University Press. Luettu 25.5.2023. <http://urn.fi/URN:ISBN:978-951-44-9249-5>.
- Rantasila, Anna. 2018. "Tahmaiset affektit. Fukushima Daiichin ydinonnettomuus YLE:n uutisoinnin verkkokommenteissa." *Lähikuva* 31 (3): 30-45. <https://doi.org/10.23994/lk.76570>.

- Rantasila, Anna. 2022. "Managing unpleasant moods: Affective discipline in Facebook discussions." *European Journal of Cultural Studies* 25 (1): 256–274. <https://doi.org/10.1177/136754942097321>.
- Rantasila, Anna, Joel Kiskola, Thomas Olsson, Aleks H. Syrjämäki, Poika Isokoski, Mirja Ilves ja Veikko Surakka. 2022. "Outlining approaches to improve online news commenting with computationally aided comment moderation." Teoksessa *Futures of Journalism: Technology driven reconfigurations in the journalism-audience relationship*, toimittaneet Ville Manninen, Mari K. Niemi ja Alan Ridge-Newman. London: Palgrave, 195–210.
- Reich, Zvi. 2011. "User Comments: The Transformation of Participatory Space." Teoksessa *Participatory journalism: Guarding open gates at online newspapers*, toimittaneet Jane B. Singer, David Domingo, Ari Heinonen, Alfred Hermida, Steve Paulussen, Thorsten Quandt, Zvi Reich ja Marina Vunjoic, 96–117. London: Wiley-Blackwell.
- Ruckenstein, Minna ja Turunen, Linda Lisa Maria. 2020. "Re-humanizing the platform: Content moderators and the logic of care." *New media @ society* 22 (6): 1026–1042. <https://doi.org/10.1177/1461444819875990>.
- Roberts, Sarah T. 2019. *Behind the Screen: Content Moderation in the Shadows of Social Media*. New Haven & London: Yale University Press.
- Santana, Arthur D. 2014. "Virtuous or Vitriolic: The Effect of Anonymity on Civility in Online Newspaper Reader Comment Boards." *Journalism Practice* 8 (1): 18–33. <https://doi.org/10.1080/17512786.2013.813194>.
- Saresma, Tuija, Reeta Pöyhkä, Aleks Knuutila, Heidi Kosonen, Marko Juutinen, Paula Haara, Urho Tulonen, Kaarina Nikunen ja Jenita Rauta. 2022. *Verkkoviha: Vihapuheen tuottajien ja levittäjien verkostot, toimintamuodot ja motiivit*. Valtioneuvoston selvitys- ja tutkimustoiminnan julkaisusarja 2022:48. Valtioneuvoston kanslia. Luettu 25.5.2023. <http://urn.fi/URN:ISBN:978-952-383-298-5>.
- Suh, Kil-Soo, Seongwon Lee, Eyung-Kyo Suh, Hoseong Lee ja Jaehoon Lee. 2018. "Online Comment Moderation Policies for Deliberative Discussion–Seed Comments and Identifiability." *Journal of the Association for Information Systems* 19 (3): 182–208. <https://doi.org/10.17705/1jais.00489>.
- Tubaro, Paola, Antonio A. Casilli ja Marion Coville. 2020. "The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence." *Big Data @ Society* 7 (1): 1–12. <https://doi.org/10.1177/2053951720919776>.
- van Dijk, Teun A. 1993. "Principles of critical discourse analysis." *Discourse and Society* 4 (2): 249–283, <https://doi.org/10.1177/0957926593004002006>.
- Wahl-Jorgensen, Karin. 2019. *Emotions, Media and Politics*. Cambridge: Polity.
- Wahl-Jorgensen, Karin. 2020. "An Emotional Turn in Journalism Studies?" *Digital Journalism*, 8 (2): 175–194, <https://doi.org/10.1080/21670811.2019.1697626>.
- Wright, Scott, Daniel Jackson ja Todd Graham. 2020. "When Journalists Go "Below the Line": Comment Spaces at The Guardian (2006-2017)." *Journalism Studies* 21 (1): 107–126. <https://doi.org/10.1080/1461670X.2019.1632733>.
- Wolfgang, David J. 2018. "Cleaning up the "Fetid Swamp"." *Digital Journalism* 6 (1): 21–40. <https://doi.org/10.1080/21670811.2017.1343090>.
- Wolfgang, David J., Stephen McConnell ja Hayley Blackburn. 2020. "Commenters as a Threat to Journalism? How Comment Moderators Perceive the Role of the Audience." *Digital Journalism* 8 (7): 925–944, <https://doi.org/10.1080/21670811.2020.1802319>.
- Wolfgang, David J., Hayley Blackburn ja Stephen McConnell. 2020. "Keepers of the comments. How comment moderators handle audience contributions." *Newspaper Research Journal* 41 (4): 433–454. <https://doi.org/10.1177/073953292096833>.